REPORT DOCUMENTATION PAGE

1. Report Date (DD MM YY) 29-06-2004

2. Report Type Book Chapter

3. DATES COVERED (from - to) N/A

4. TITLE AND SUBTITLE Celebrating the usefulness of pictorial information in visual perception

5a. Contract Number:
5b. Grant Number:
5c. Program Element:
5d. Project Number:
5e. Task Number:
5f. Work Unit Number: 60283

6. AUTHORS
Jeremy Beer

7. PERFORMING ORGANIZATION NAME AND ADDRESS
Naval Health Research Center DET, Directed Energy Bioeffects Laboratory
8315 Navy Road
Brooks City Base TX 78235

9 PERFORMING ORGANIZATION REPORT NUMBER

8. SPONSORING/MONITORING AGENCY NAME AND ADDRESS
Chief, Bureau of Medicine and Surgery
MED-02
2300 E St NW
Washington DC 20372-5300

10. Sponsor/Monitor's Acronyms(s) BuMed

11. Sponsor/Monitor's Report Number

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT (maximum 200 words)

This book is a collection of biographical essays describing the influence of Julian Hochberg, a leading researcher in vision science and human performance modeling. In this chapter, Jeremy Beer, who was Hochberg's doctoral student, describes three areas in which Hochberg's experimental approach remains influential in vision research. The first area comprises the comparison of motion information vs. pictorial depth cues (e.g. linear perspective, relative size) in moving viewers' perception of distance and slant. Hochberg demonstrated that pictorial cues can overcome other sources of information to determine how viewers will perceive a three-dimensional scene. These principles continue in the design of modern cockpit displays incorporating features like "highway in the sky". The second area comprises the comparison of motion information vs. pictorial cues in the perception of time-to-collision. As in the first topic, recent work has demonstrated that pictorial cues can dominate other kinds of information in important visually controlled tasks such as vehicle braking and interceptive action. The third area comprises the effects of display boundaries on the perception of extended scenes; recent work in this field has determined that human operators perceive scenes according to a distorted geometry when the boundaries of the display are restricted.

20040809 108

15. SUBJECT TERMS
Visual Perception, Motion Perception, Depth Cues, Visual Performance

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	b. THIS PAGE			NHRC DET DEBL OIC, CDR, MSC, USN
UNCL	UNCL	UNCL	UU		19b. TELEPHONE NUMBER (INCLUDING AREA CODE) COMM/DSN: (210) 536-6924/240-6924

Celebrating the usefulness of pictorial information in visual perception

Jeremy Beer

Naval Health Research Center Detachment DEBL, and the Henry M. Jackson Foundation for the Advancement of Military Medicine, Brooks City-Base, Texas

I encountered Julian Hochberg's constructivist approach to perception as his doctoral student at Columbia University. In the years since, this approach has continued to influence the design of my experiments, especially those examining the moving viewer's perception of scenes and events, and to guide my thinking about perception in general. One of its particular strengths is that it has transcended polemic, not because it lacked unifying principles and strong opinions (indeed, some of these principles and opinions have provoked heated debate), but rather because the approach was so purely empirical. In my recollection, whenever Hochberg encountered a conflict between two theories of perception, he would work swiftly to articulate an unambiguous prediction from each, and then craft an experimental test to determine which would prevail. In devising this empirical test, he would always pose a specific question regarding the perceptual component of interest; importantly, this question was usually not "Can the viewer use this class of information?", but rather, the more probing "*Does* the viewer use this class of information even when other sources are available?" Observing this rigorous empirical approach in Hochberg's laboratory taught me to be careful about wielding a rigid and universal theoretical hammer to

attack questions in perception.[1] The experiments I performed with him taught me that the factors underlying a perceptual competence can change in the presence of stimulus transformations, and also that tolerance of such transformations is engineered into some of these underlying perceptual components.

Hochberg's view of perception has remained particularly influential to me in three areas. These include the respective roles of motion vs. pictorial information in the perception of three-dimensional configurations and events; the similar conflict between dynamic and pictorial information in the judgment of time-to-contact; and the effects of display boundaries on the dimensions of space perceived by the viewer.

Ames Phenomena: Pictorial Cues vs. Motion in Depth and Event Perception

Throughout my time at Columbia, a squadron of Ames Windows and Ames-derived objects stood prominently in the Hochberg lab, defying bystanders to ignore their persistent rubberiness. These devices hold a personal significance for me because I remember everyone in the lab discussing them, playing with them, and feeling taunted by them over the years, in spite of the fact that we were all continually and variously engaged in any number of other projects.

The Ames object is a trapezoidal contour with converging edges and shading cues painted on both faces, which induces the illusion of a rectangular window slanting into depth. Because of the pictorial depth cues of linear perspective and relative size (whereby objects that subtend lesser visual angles are perceived as farther away than objects that subtend greater visual angles), Ames

windows are almost always perceived as if the shorter edge is more distant, even when it actually juts forward. This leads to the classic Ames effect in which the trapezoid is rotated about a vertical axis, and the viewer instead perceives a window yawing back and forth in oscillation.

We employed the Ames objects in a sequence of experiments and demonstrations, all of which were constructed to pit the pictorial depth cues against motion information. The pictorial cues are misleading much of the time (*viz.*, whenever the short edge is not farther away). In contrast, the motion information should specify the object's actual layout and slant, if viewers are capable of extracting a rigid configuration from the dynamic image transformations that occur during the object's, or their own actively-initiated movement.

In spite of this, none of our experiments was successful in reliably banishing the Ames objects' tendency to induce motion and depth illusions when viewed from any vantage point other than directly above, or within a short distance. Not content with the classic Ames demonstration in which the continuously-rotating trapezoid appears to swing back and forth, we tried using an entertaining variety of devices to defeat the painter's cues. We replaced the painted shading with texture patterns such as uniformly spaced dots, which increased the information specifying the trapezoid's flatness and introduced an optical motion gradient that was unbiased by false illumination cues; nevertheless, the perspective of the converging edges and the relative size of the vertical edges prevailed, and the rotating object still appeared to swing like a screen door. We

pierced the window with a solidly-mounted metal rod, and it still appeared to swing, now as an impossible figure with an apparently-flexible bar repeatedly violating the continuity of its solid parts. We constructed a new figure comprising two Ames trapezoids back to back, and placed this rigid, planar, hexagonal figure in a yawing oscillation (an actual movement that resembled the illusion of yawing oscillation described above), and it appeared to crease along its central spine like a butterfly. Finally, we froze the original trapezoid in a fixed orientation with the short edge in front, and had viewers generate motion by swaying their own vantage point from side to side. These movements made the stationary object appear to swing, because the perception of the trapezoid's optical deformation was coupled with the illusory, reversed perception of its slant: Assuming the depth cues' accuracy, the opening and closing of the object's image could be explained only if the window were yawing in synchrony with the viewer's movements. The fact that the cues were inaccurate does not alter their strength, nor their coupling with the viewer's interpretation of optical deformation. Throughout these situations, it became increasingly difficult to dismiss the pictorial cues as artifacts of a painted world, because they worked so effectively against motion information specifying the actual, rigid distal configuration. (See Hochberg, 1986, for a more detailed description of the conditions under which the Ames effects are observed).

Augmenting visual displays with depth information has remained a fertile topic of inquiry. Van den Berg (1994) and Vishton, Nijhawan, and Cutting (1994) claimed that adding veridical depth information enhances heading

perception during self-motion, although Ehrlich et al. (1998) reported subsequently that this addition is not useful without an appropriate extraretinal ? eye movement signal. Adding depth information to optic flow patterns reportedly increases MST neurons' heading selectivity and sensitivity during ocular pursuit (Upadhyay, Page, & Duffy, 2000). In addition, the enhancement or addition of pictorial depth cues has been shown to influence the effectiveness of vehicle displays. Some years ago, I built up some virtual clouds to introduce illusory depth cues in a synthetic flight environment, and found that these objects were capable of distorting a pilot's judgment of the aircraft sink rate in a landing approach task (Beer et al., 1998). And new-generation "pathway in the sky" aviation displays are designed specifically to add veridical perspective and relative size cues to the pilot's visual environment (Snow et al., 1999). My interest in these pictorial cues continues unabated, and some of my most vivid recollections about titrating visual depth information empirically remain those of our playful efforts to make the Ames Window stand up for itself and look like an unyielding object in the Hochberg lab.

Relative Size vs. Motion in Time-to-Contact Judgment

The second area in which Hochberg's emphasis on pictorial depth information has proven influential in perception is that of time-to-contact judgments. Patricia DeLucia has produced a notable body of work in this area, which includes findings relevant to self-motion control, collision avoidance, and interceptive action. Like the Ames investigations, these studies juxtaposed optical motion

information (which could, in theory, specify irrefutably the time remaining until a viewer will contact an approaching or approached object) and pictorial depth cues (which might be configured to alter or contradict the optically specified solution). In these experiments, DeLucia constructed environments in which the raw expansion information specified one perceived configuration while the pictorial cue of relative size could specify a contradictory solution. This line of research continued Hochberg's tradition of articulating conflicting predictions from competing theories clearly, and then testing the predictions unambiguously.

An object's optical expansion can specify the time remaining until the object reaches the observer or *vice versa* (Lee, 1976); if the expansion remains above threshold, this information source is largely independent of the object's size. But according to the relative size cue, larger images typically belong to nearer objects (see above); for this reason, an observer approaching two objects that subtend different visual angles will expect to reach the larger object first, because it looks nearer. DeLucia first effected the competitive comparison between optical expansion and relative size in a paradigm that required the viewer to judge which of two approaching objects would arrive first (DeLucia, 1991). Large, distant objects were judged reliably to arrive at the viewer's position before small, near objects that would actually have arrived sooner. In a subsequent study, DeLucia (1994) instructed subjects (who were controlling their movement in a visual self-motion simulation) to approach a fixed object as closely as possible and then jump over it without colliding. As was the case with the approaching objects, the landmarks' projected size influenced control

movements consistently, with subjects jumping earlier to clear large objects than to clear small objects they were approaching at equal speeds. These two sets of findings indicate that predictive models of distance perception and self-motion control must include the effects of pictorial information, particularly relative size.

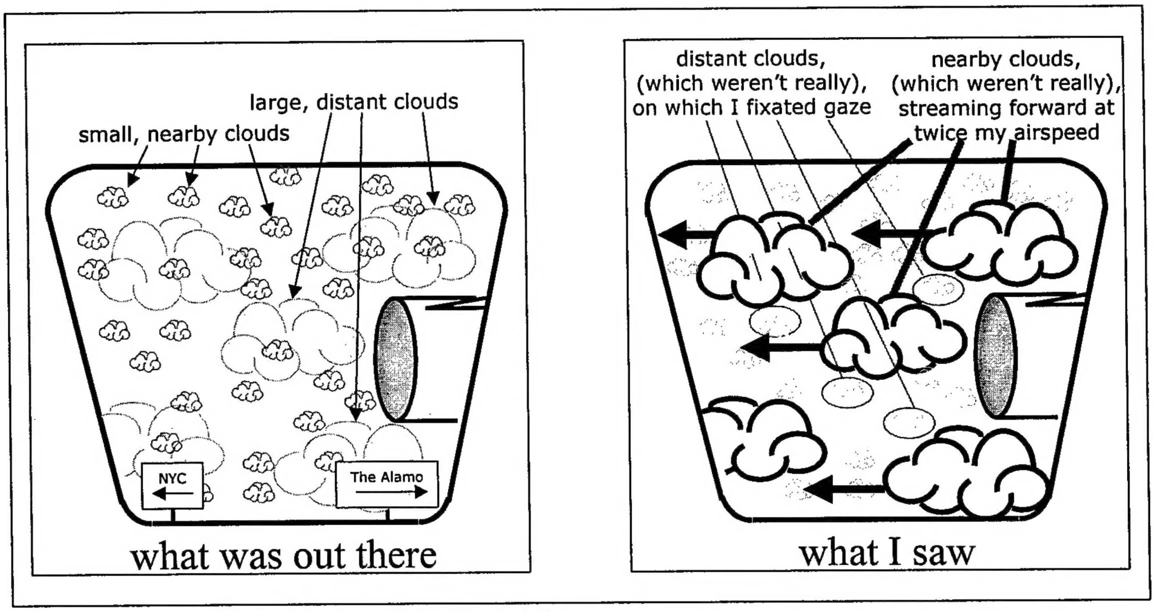

Figure 1.

It is worth noting that the interaction between visual motion and pictorial depth information, explored in these first two research areas, can cause unexpected perceptual consequences, particularly in observers viewing unfamiliar scenes. Once, while flying in a plane at high altitude, I looked down through a fine-grained layer of high clouds to a coarser layer of lower, larger clouds, and experienced the shocking and persistent perception that the large clouds (which looked nearer, but weren't) were blasting forward at twice the speed of the aircraft. After some head-scratching, I managed to reconcile this perception with my disbelief in 1000-knot jetstreams, by considering that relative size can alter the perception of distance in optic flow environments: I must have been fixating

the small, near clouds, seeing them as more distant, and then misinterpreting the motion parallax caused by the large, far clouds streaming in my retinal field, in the same direction as my own travel (Figure 1).[2]

Effects of Display Boundaries on the Perception of Extended Scenes

The third way in which Hochberg has been profoundly influential is through his emphasis on how a scene is typically perceived across a succession of views, and on how this perception can be affected by the geometric boundaries governing the successive views. Shortly before I completed my doctorate, he expressed this emphasis forcefully and eloquently in a conversation regarding the proposed dichotomy between "what" vs. "where" processing streams in the brain (Ungerleider & Mishkin, 1982). Hochberg was clearly uncomfortable with the possibility that this dichotomy could be over-interpreted and adopted as dogma, in the face of evidence indicating that the divergence between the two classes of information is not absolute. I remember particularly his pointing out that sometimes a perception of the "what" kind is ambiguous or impossible unless and until the viewer manages to integrate information across a succession of "where" perceptions. In one demonstration of this principle, he displayed successive close-up views through a round aperture of the individual corners of a cross-shaped figure, which was much larger than the aperture (Hochberg, 1986). The sequence of partial still views was unintelligible, looking like a disjointed set of pictures of a clock face, unless some integrating structure was provided. One way to convey this structure was to present a prior "long shot" view of the entire

object as seen from afar. Alternatively, the partial views could be tied together by moving the object behind the aperture to reveal its features over time. In the latter case, the perception of global shape depended on viewers' ability to integrate visual motion across time and thereby build up a defining group of "where" relationships among the object's components.

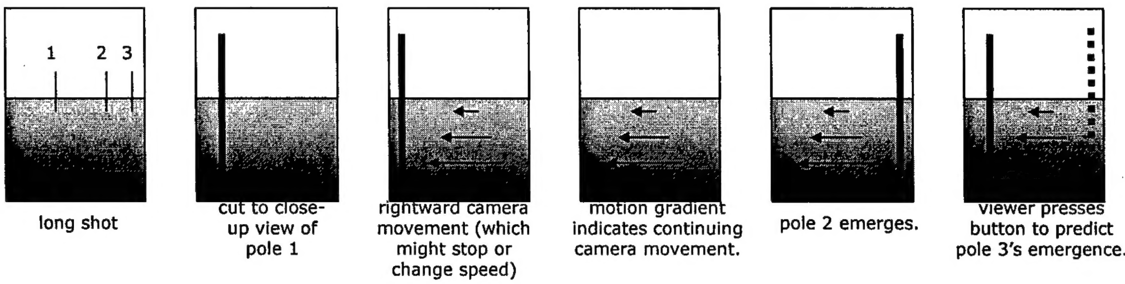

Figure 2.

This building-up of a spatial percept over time and across views comprised the foundation for my dissertation research, which examined the metric of the extended space that viewers can perceive when a viewing aperture (or a movie camera, or the viewer's limited instantaneous field of view) moves relative to the figures or landmarks in a scene. Examples of this perceptual competence include a driver's ability to maintain spatial awareness of other cars on the road as they move into and out of view (e.g. from the windshield to the rear-view mirror), and also the moviegoer's ability to understand the layout of a room depicted by a moving camera even when the room is never shown in its entirety.[3]

In a series of experiments, we used chronometric modeling to map the extended spaces viewers perceived while observing simulated self-motion displays in which the camera tracked laterally (Beer, 1993). The viewer's task

was to press a button during the camera movement to predict the emergence on-screen of a widely-displaced peripheral target landmark, whose position in the scene had been shown in a prior "long shot", or panoramic view (Figure 2). The experiments identified two characteristics of viewers' ability to perceive the dimensions of an extended scene configuration as revealed by a moving camera. First, viewers were able to integrate optic flow over time; specifically, they perceived their depicted self-motion fairly accurately as the integral of camera speed over time (including changes in speed and pauses in the movement), up to a limiting boundary. Within this boundary, it was determined that when the camera moved, viewers could predict the emergence of the target landmark at close to the ideal response time. This ideal time corresponded with the span of a camera movement that should be required to reveal the target, as specified in the prior view: The wider the lateral spacing of the target, the later the response, *up to the limiting boundary*.

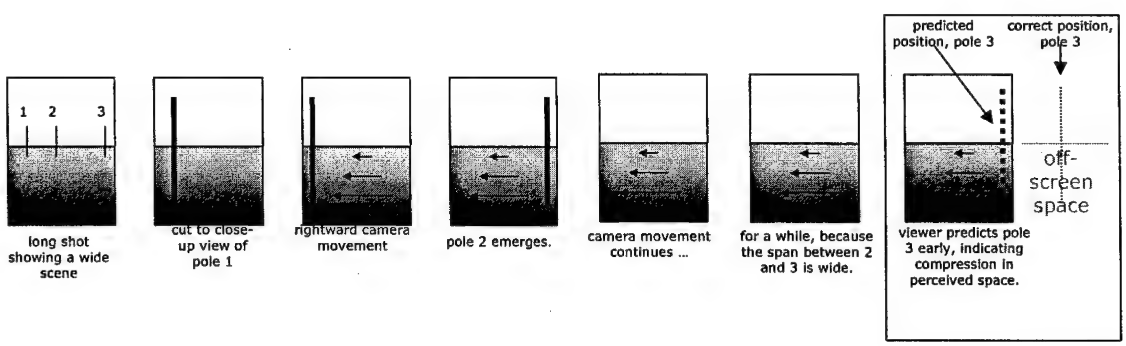

Figure 3.

The second characteristic identified in these experiments comprised this limiting boundary, beyond which the geometry of the perceived space defined by the button-presses changed. When the prior panoramic view displayed a scene configuration so large that viewers must integrate the lateral optic flow across an imagined span that was wider than the close-up view could display at one time, systematic distortions emerged in the space perceived beyond the edges of the screen. In particular, while the timing of the prediction responses continued to lengthen linearly with target distance when these very wide scenes were displayed (as it should if viewers were retaining information accurately from the prior view and using it to integrate the subsequent camera motion information), the slope of the response-time curve flattened. Viewers were compressing these scenes perceptually and predicting the target landmark's emergence early, and the more the imagined span exceeded the width of the close-up view, the greater the scene compression became (Figure 3). This compression effect indicated that while viewers are capable of using remembered information in conjunction with optic flow to perceive and generate expectations about scenes extending beyond the edges of the display, there are boundaries beyond which this perception departs from a Euclidean metric. Nevertheless, it remains true that to the extent a remembered geographic configuration comprises a "what" representation, generating it by integrating motion information among a succession of partial views constitutes a perceptual building-up among "where" representations, just as Hochberg suggested.

Conclusion

In these and in other areas of inquiry, Hochberg's approach to perception has influenced me meaningfully, as it has influenced the fields of vision science, human engineering, and film theory. Had I not stumbled into a teaching assignment with him 17 years ago, I would not have been drawn in by his enthusiasm for projective geometry, by his rigorous emphasis on the importance (and the limitations) of optic flow, and by his unflagging celebration of pictorial information in perception. This celebration has enriched my understanding and experience, because thinking about perception is its own reward, a reward that is particularly satisfying when one is exploring among the landmarks, textures, monuments, shadows, vehicles, and figures of a novel environment.

References

Beer, J. M. A. (1993). Perceiving scene layout through an aperture during visually simulated self-motion. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 1066-1081.

Beer, J. M. A., Ghani, N., Previc, F. H., & Campbell, J. M. (1998). Speed can influence aim-point judgments even before illusory depth cues enter the picture. *Perception, 27, Supplement: Proceedings of the European Conference on Visual Perception*, 58.

van den Berg, A. V., & Brenner, E. (1994). Humans combine the optic flow with static depth cues for robust perception of heading. *Vision Research, 34*, 2153-2167.

Cutting, J. E. (1986). *Perception with an Eye for Motion.* Cambridge, MA: MIT Press.

Cutting, J. E., Springer, K., Braren, P. A., & Johnson, S. H. (1992). Wayfinding on foot from information in retinal, not optical, flow. *Journal of Experimental Psychology: General, 121*, 41-72.

DeLucia, P. R. (1991). Pictorial and motion-based information for depth perception. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 738-748.

DeLucia, P. R. (1994). Pictorial and motion-based depth information during active control of self-motion: Size-arrival effects on collision avoidance. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 783-798.

Ehrlich, S. M., Beck, D. M., Crowell, J. A., Freeman, T. C., & Banks, M. S. (1998). Depth information and perceived self-motion during simulated gaze rotations. *Vision Research, 38*, 3129-3145.

Hochberg, J. (1986). Visual perception of real and represented objects and events. In N. J. Smelser & D. R. Gerstein (Eds.), *Behavioral and Social Science: Fifty Years of Discovery.* Washington, DC: National Academy Press.

Lee, D. N. (1976). A theory of visual control of braking based on information about time-to-collision. *Perception, 5*, 437-459.

Snow, M. P., Reising, J. M., Liggett, K. K., & Barry, T. P. (1999). Flying complex approaches using a head-up display. *Proceedings of the 10th International Aviation Psychology Symposium.* Columbus, OH: Ohio State University.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. Ingle, M. Goodale, and R. Mansfield (Eds.), *Analysis of Visual Behavior* (pp.549-586). Cambridge, MA: MIT Press.

Upadhyay, U. D., Page, W. K., & Duffy, C. J. (2000). MST responses to pursuit across optic flow with motion parallax. *Journal of Neurophysiology, 84*, 818-826.

Vishton, P. M., Nijhawan, R., & Cutting, J. E. (1994). Moving observers utilize static depth cues in determining their direction of motion. *Investigative Ophthalmology & Visual Science, 35 (Supplement)*, 2000 (Paper #3444).

Contact information

Jeremy Beer

Naval Health Research Center/DET DEBL, and the Henry M. Jackson Foundation for the Advancement of Military Medicine

8315 Navy Road

Brooks City-Base, TX 78235

Email: jeremy.beer@navy.brooks.af.mil

---

[1] Minimum principles, assertions of direct perception, and computational reverse-projection algorithms are examples of theoretical tools that turn brittle when it is demonstrated that viewers tolerate certain inconsistencies in a visual display. These inconsistencies include the coexistence of mutually contradictory spatial information, impossible geometric transformations, and the depiction of non-rigidity in the structure of distal objects.

[2] When an observer moves through the world and fixates an object located apart from the direction of locomotion, nearer objects will typically stream away from the aimpoint in the retinal field, as more distant objects stream towards it, in the same direction as the viewer's movement (Cutting, 1986; Cutting et al., 1992).

[3] This ability to perceive extended spaces behind and beyond the edges of the screen is demonstrated clearly when one is given the opportunity to explore an actual scene that has been viewed previously in a cinematic sequence or a computer-generated graphic rendering. With the advent of virtual architectural tours and simulated mission rehearsals, the commercial and operational

application of this perceptual competence is becoming commonplace. Its power remains striking, however, as I discovered a few months ago at a diner near Barstow, California, which I had seen previously in the strange and atmospheric film "Bagdad Café". As I entered the store, I was familiar with its configuration; I knew in a relative sense where the tables and counter and adjoining rooms would be (though according to the research described above, I might not have known exactly how many steps would be required to move from one of these features to another).